

(11)Publication number : 09-081567  
(43)Date of publication of application : 28.03.1997

**G06F 17/28**

(72)Inventor : SHIMAZU MIWAKO  
KUMANO AKIRA

[illegible]

**BEST AVAILABLE COPY**

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-81567

(43) 公開日 平成9年(1997)3月28日

(51) IntCl.<sup>8</sup>

G 0 6 F 17/28

識別記号

庁内整理番号

F I

G 0 6 F 15/38

技術表示箇所

R

審査請求 未請求 請求項の数 2 O L (全 13 頁)

(21) 出願番号 特願平7-235632

(22) 出願日 平成7年(1995)9月13日

(71) 出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72) 発明者 島津 美和子

神奈川県川崎市幸区小向東芝町1番地 株式会社東芝研究開発センター内

(72) 発明者 熊野 明

神奈川県川崎市幸区小向東芝町1番地 株式会社東芝研究開発センター内

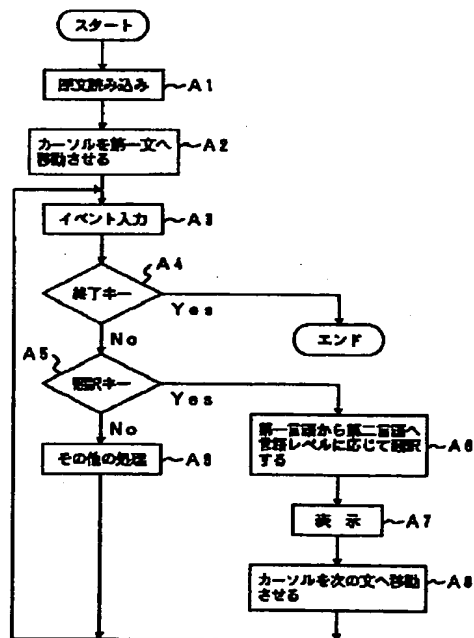
(74) 代理人 弁理士 鈴江 武彦

(54) 【発明の名称】 翻訳方法

(57) 【要約】

【課題】 言語能力とは関係なく第一言語から第二言語に翻訳された文章を容易に内容把握できるようにする。

【解決手段】 第一言語の文章を第二言語の文章に翻訳する翻訳方法であって、第二言語の文章の文章形態を特定するための言語レベルを入力し、第一言語の文章を、入力された言語レベルに応じた文章形態による第二言語の文章に翻訳し、この翻訳によって得られた第二言語の文章を提示する。



## 【特許請求の範囲】

【請求項1】 第一言語の文章を第二言語の文章に翻訳する翻訳方法であって、

第二言語の文章の文章形態を特定するための言語レベルを入力し、

第一言語の文章を、入力された言語レベルに応じた文章形態による第二言語の文章に翻訳し、

この翻訳によって得られた第二言語の文章を提示することを特徴とする機械翻訳方法。

【請求項2】 第一言語の文章を第二言語の文章に翻訳する翻訳方法であって、

第一言語の文章を第二言語の文章に翻訳し、

第二言語の文章の文章形態を特定するための言語レベルを入力し、

翻訳によって得られた第二言語の文章を、入力された言語レベルに応じた文章形態による第二言語の文章に変換し、

この変換によって得られた第二言語の文章を提示することを特徴とする機械翻訳方法。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、自然言語処理技術に関わり、より詳しくは、自然言語による文章を他の言語の文章に翻訳する翻訳方法に関する。

## 【0002】

【従来の技術】従来から多くの機械翻訳システムが開発されているが、それらは主として企業内の、英語をある程度学んだ成人によって利用されてきた。しかし、翻訳システムの低価格化がインターネットの浸透と結び付き、翻訳システムが個人でも利用されるようになってきた。

【0003】近年、外国語を知らない子供も、自習で外国語を学ぶ手段として、その言語を母国語とする人と通信したり、その言語の文書を読む必要がでてきており、そのために翻訳システムが必要になっている。しかし、現状では、一通りの読み書きの能力を身に付けていない学童前の幼児や小学生低学年を想定した訳文を作成する翻訳システムはない。つまり、第一言語が幼児向けの文章であっても現在のシステムでは大人の文法で大人の文章を訳出している。

【0004】まず、子供の文章の代表的な特徴を掴んでおく。次は、小学1・2年生向けの文章である。「藤六は、おおきな よめいりだんすを、せなかに しょって、やっと、となりむらとの むらざかいの もりまで、あるいていきました。そらはすっかり はれわたって、けさの あの ふきぶりは、まるで ゆめだったとでも いいたいような、いい てんきに なりました。」(ただし、漢字「藤六」には「とうろく」のルビが付与される)。

【0005】以上の文章で見られる特徴を挙げると次の

ようになる。大部分をひらがなで記述する、未習だがどうしても漢字で表示する必要がある語りは仮名を付ける、ひらがなが多いため文節毎に分ち書きする、読点が多い、「である」調よりは「ですます」調が多い、である。

【0006】さらに、小学生に関しては文部省の「小学校学習指導要領」に学年別漢字配当表(計1006文字)が掲載されているので、漢字の使用率はこれに準じていることが多い。

【0007】もう少し年齢が上がると、分ち書きはしなくなるが、どの漢字のどの用法(読み方)をどの学年で学ぶという知識を持っていなければ子供向けの文章は書けない。つまり、原則として既出用法は漢字で書けるが、それ以外で漢字を用いるときは必ず振り仮名を振る。

【0008】次は、小学5、6年生向けの文章である。「年月がたって、ヤーコは、りっぱに大きくなった。年ごとにちからがついて、おなじ年の友だちととくみあいをすれば、かならずけがをさせた。ちからがつよくて、あばれ牛のつのをつかみ、くびをねじふせておさこんでしまうくらいだった。」(ただし、漢字「年月」には「としつき」、「年」には「とし」、「牛」には「うし」のルビが付与される)。

【0009】この文章は大人であれば、「年月がたってヤーコは立派に大きくなった。年ごとに力がついて同じ年の友達と取組み合いをすれば、必ず怪我をさせた。力が強くて、暴れ牛の角を掴み、首を握り伏せて押さえ込んでしまうくらいだった。」といったようになったであろうと思われる。

【0010】中学・高校では小学校ほど厳密ではないが、それぞれ「常用漢字の大体を読み、小学校の学年別漢字配当表の漢字を主として1000字程度の常用漢字について使い慣れること」、「常用漢字の読みになれ、主な常用漢字が書けるようになること」と目標設定されている。

【0011】このように見てくると子供向けの訳文を作るには特別な配慮が必要であることが分かる。例えば、次のような英文があったとする。“The Clintons have decided to go West for their summer vacation this year. After spending their last two summer vacations on the Massachusetts island. White House sources say the First Family will head to the Grand Teton mountains in Wyo. for a week or more of golf, horseback riding and buffalo barley soup at the home of Democratic Sen. John D. Rockefeller IV. Then they're off to Hawaii for a few days off R&R.”この英文の訳文として、大人が読者の場合、「クリントン一家は今年の夏期休暇には西に行く」と決定した。この2年の夏期休暇はマサチューセッツの島で過ごしたため、大統領一家はワイオミング州のグラン

ドティートンの山々へ向かい、民主党のジョン・D・ロックフェローIV上院議員の家で一週間程度ゴルフや乗馬やバッファローの大麦スープを満喫する予定である、とホワイトハウス関係者は言っている。その後、彼らは2、3日の保養休暇のためにハワイに向かう。」といったようになると考えられる。

【0012】また中学生であれば、「クリントン一家は今年の夏の休暇には西に行くことを決めた。ここ2年の夏の休暇はマサチューセッツの島で過ごしたため、大統領一家はワイオミング州のグランドティートンの山々へ向かい、民主党のジョン・D・ロックフェローIV上院議員の家で一週間ほどゴルフや乗馬やバッファローの大麦スープを楽しむ予定である、とホワイトハウスの関係者達は言っている。その後、彼らは2、3日の保養休暇のためにハワイに向かう。」といったようになる。(ただし、漢字「休暇」には「きゅうか」、「大統領」には「だいとうりょう」、「上院議員」には「じょういんぎいん」、「保養休暇」には「ほようきゅうか」のルビがそれぞれ付与される)。

【0013】一方、読み手が小学生の場合、難しい単語には易しい言い回しに変更したり、難解な読みには振り仮名をふったりする必要が出てくる。さらに、年齢が下の小学生低学年の場合、大幅な変更が必要になる。具体的には用いる語をかなり制限し、ひらがなを多くし、文節毎に分かち書きし、読点も多用するといったことである。場合によっては訳出を省いたり、語を補ったりする工夫や、文章も「です・ます」調を用い、語りかけるような文体にするなどの変更が必要となってくるであろう。

【0014】この場合、先の英文に対しては次のような日本語訳文が考えられる。「クリントンいっかは、ことしのなつ休みには 西に行くことをきめました。おとしときょねんのなつ休みは マサチューセッツのしまですごしたため、だいとうりょういっかは ことしは ワイオミングしゅうグランドティートンの山へいき、ジョン・D・ロックフェローさんの家で ーしゅうかんくらいゴルフや馬のりやバッファローのスープをたのしみます、とホワイトハウスのひとたちは、いってます。そのあと、だいとうりょういっかは2、3日のお休みにハワイにいきます。」(ただし、漢字「家」には「いえ」、「馬」には「うま」のルビが付与される)。

【0015】以上のように、利用者にとって最も分かりやすい訳文を生成するには、利用者の母国語の言語能力に応じて訳文の生成や加工を行ったり、対象の学年で決まっている用法と記述様式に基づいて訳文を生成するシステムが必要となる。

【0016】

【発明が解決しようとする課題】このように従来の機械翻訳システムでは、第一言語の文章を訳出する際には、一律に一般の成人向けの文章を訳出しており、一般の成

人レベルに達していない読み手を想定して、あらかじめ訳文を生成する、あるいはその成人向けの訳文をさらに加工するということを行われていなかった。

【0017】そのため、読み手は訳文が母国語でありながら、提示された文章を理解することができず、翻訳が第一に目指す目標を果たしているとは言えなかった。もとの文章がたとえ大人の文章であっても、このように読み手の言語能力に合わせて訳文を生成・調整する必要性は大きい。しかし、現状では原文書が子供向けの易しい文章であっても、訳文が大人向けの難解な文章となってしまうことも少なくなかった。

【0018】本発明は前記のような事情を考慮してなされたもので、言語能力とは関係なく第一言語から第二言語に翻訳された文章を容易に内容把握できるようにすることが可能な翻訳方法を提供することを目的とする。

【0019】

【課題を解決するために手段】本発明は、第一言語の文章を第二言語の文章に翻訳する翻訳方法であって、第二言語の文章の文章形態を特定するための言語レベルを入力し、第一言語の文章を、入力された言語レベルに応じた文章形態による第二言語の文章に翻訳し、この翻訳によって得られた第二言語の文章を提示することを特徴とする。

【0020】また本発明は、第一言語の文章を第二言語の文章に翻訳する翻訳方法であって、第一言語の文章を第二言語の文章に翻訳し、第二言語の文章の文章形態を特定するための言語レベルを入力し、翻訳によって得られた第二言語の文章を、入力された言語レベルに応じた文章形態による第二言語の文章に変換し、この変換によって得られた第二言語の文章を提示することを特徴とする。

【0021】このような方法によれば、利用者の言語レベルに応じて利用者が通常接している文章形態で原文に対する訳文が提示されるので、母国語以外の文章を人手を借りることなく理解し、情報を得ることができる人々の層が増加する。

【0022】

【発明の実施の形態】以下、図面を参照して本発明の実施の形態について説明する。

(第1実施形態)図1は第1実施形態に係わる機械翻訳装置の構成を示すブロック図である。図1に示すように、機械翻訳装置は、入力部101、システム制御部102、編集部103、翻訳部104、翻訳辞書部105、表示部106、及び文書記憶部107によって構成されている。

【0023】入力部101は、キーボードやマウス、OCR、磁気テープ装置、磁気フロッピーディスク装置などから構成される。入力部101は、翻訳処理の対象となる文章(例えば、日本語文、英文、中国語文)のデータを入力するのを始め、各種編集・修正等のイベントを

実行させるためのコマンドを入力するためのものである。例えば、入力部101として設けられたキーボードには、処理の終了を指示するための「終了キー」、翻訳の実行を指示するための「翻訳キー」等が設けられている。

【0024】システム制御部102は、本装置全体の制御を司るものである。システム制御部102は、入力部101からの文章データや、辞書データを各部に送ったり、入力部101から入力されたコマンドの指示に基づいて翻訳すべき文章（例えば一文単位）の選択、辞書の選択などを行なうものである。

【0025】編集部103は、システム制御部102で選択された第一言語の文章データ等に対する編集の他、表示部106の表示画面に対する編集、すなわち入力部101から入力された指示に応じた表示画面上でカーソルの移動、言語レベル入力用の画面の設定、翻訳部104における翻訳処理によって得られた訳文の表示部106での表示等の制御を行なうものである。

【0026】翻訳部104は、システム制御部102で選択された第一言語による文章データに対して、システム制御部102で選択された翻訳辞書部105の辞書を参照しながら翻訳処理を行なう。

【0027】翻訳辞書部105は、翻訳部104が翻訳処理を行なう際に使用する各種辞書データを収容している。翻訳辞書部105に含まれる辞書の内容については後述する。

【0028】表示部106は、入力部101から入力された翻訳対象とする第一言語の文章（原文）や、翻訳部104により訳出された第二言語の文章（訳文）の他、言語レベル入力用の画面等を表示するものである。表示部106は、文字表示が可能なものとして、通常CRTや液晶ディスプレイ等が使われる。

【0029】文書記憶部107は、第一言語の文章データと第二言語の文章データとをそれぞれ独立して記憶すると共に、第一言語データとその翻訳結果としての第二言語のデータを対応付けて記憶することができる。

【0030】次に、翻訳辞書部105に含まれる辞書の内容について説明する。翻訳辞書部105は、第一言語から第二言語への読み手の言語レベルに合った翻訳を行なうための辞書であり、言語レベルに関係なく共通の知識を記憶した辞書と、言語レベルごとに異なる知識を記憶した辞書の二段階構造になっている。

【0031】図1に示すように、言語レベル共通の知識を記憶する辞書には、語尾等に変化のある単語・熟語をその原形に変換するための活用変化辞書105a、第一言語を解析するための文法が記憶された解析文法辞書105b等が含まれている。一方、言語レベルごとに異なる知識を記憶した辞書には、第一言語の単語・熟語に対応する第二言語の訳語がレベル別に分類され、その品詞情報と共に記憶されるレベル別単語・熟語辞書105

c、利用者の言語レベルに適した変換情報が記憶されたレベル別変換文法辞書105d、利用者向けに書き換えられた文の構造を決定するレベル別生成文法辞書105e、さらに語尾等の語形を変化させて利用者向けの翻訳文を完成させるレベル別形態素生成文法辞書105f、言語レベルに応じた特殊な書式に変換するために必要な書式知識辞書105g等を収容したものである。各辞書の内部データの詳細については後述する。

【0032】次に、第1実施形態における機械翻訳装置の動作について説明する。入力部101から入力された文章に対する翻訳処理の流れを、第一言語の英語による原文を、第二言語の日本語による訳文に翻訳する場合を例にして、図2のフローチャートを参照しながら説明する。

【0033】ここでは、具体例として、表示部106によって図3に示すように表示された英語の文章のうち、第5行目からの「Construction has begun on its three arenas, including the Olympic stadium, which after the Games will be trimmed by one third and converted into a major-league baseball park.」という文に対する翻訳の要求を受け、翻訳・表示する場合を例にして説明する。

【0034】まず、入力部101（キーボードやOCR等）から第一言語の英文による原文（文章データ）が読み込まれる。システム制御部102は、入力された第一言語の原文の文章データを、文書記憶部107に記憶させる（ステップA1）。

【0035】システム制御部102によって原文（文章データ）が保存されると、編集部103は、システム制御部102を介して得られる文書記憶部107に記憶された文章データをもとに、図3に示すように、表示部106によって表示画面206中に原文を表示させる。また、編集部103は、翻訳対象とする文を指定するためのカーソルを第一文に移動させる（ステップA2）。

【0036】以下、入力部101から入力されるイベントの実行を指示するコマンドに応じて処理を実行する（ステップA3）。ステップA3において、処理の終了または翻訳実行の要求以外のイベントが入力された場合、ステップA9において入力された指示に応じたイベントの処理を行なう。これには、例えばカーソルの移動が含まれる。

【0037】カーソルの移動が指示されると、編集部103は、指示に応じて表示画面上のカーソルを移動させる。本実施形態では、表示画面上でカーソルが位置する1文、例えばカーソルが先頭にある一文を、翻訳のイベントが指示された際の翻訳対象とする。カーソルの移動は、例えば原文を翻訳しようとする際に、現在のカーソル位置の直後の文ではなく、その前や後にあるその他の文を翻訳したいときなどに、対象とする文を指定するために行なわれる。

【0038】また、入力部101のキーボードに設けられた「終了キー」が押され、処理の終了が指示されると、システム制御部102は、処理を終了させる(ステップA4)。

【0039】一方、「翻訳キー」が押された場合(ステップA5)、システム制御部102は、翻訳要求が出されたと見なし、以下のステップA6～A8の処理を各部において実行させる。まず、システム制御部102は、翻訳部104に対して、第一言語の原文を第二言語の訳文に翻訳する翻訳処理を、言語レベルに応じて実行させる。すなわち、翻訳処理によって生成される第二言語の訳文を、言語レベルによって特定される表示形態となるようにする。

【0040】ここで、第1実施形態における機械翻訳装置では、利用者の言語レベルが一般の成人のレベルに達しておらず、言語レベルに応じた翻訳処理が必要な場合には、例えば図4(a)～(c)に示すような、言語レベル入力用の画面を用いて言語レベルを指示し、その言語レベルによる翻訳を実行させることができる。

【0041】言語レベルの入力としては、具体的にはその利用者の年齢や学年や最終学歴といった項目が用いられる。図4(a)に示す言語レベル入力用の画面206aは、利用者の年齢に基づいて言語レベルを入力するためのものである。図4(b)(c)に示す言語レベル入力用の画面206b、206cは、利用者の学歴に基づいて言語レベルを入力するためのものである。また、言語レベル入力用の画面206a、206cでは、選択肢がボタンで表示されており、該当するボタンを選択することによって言語レベルが入力される。また、言語レベル入力用の画面206cでは、利用者学歴が直線上の位置で示されており、この直線上でカーソルを移動させ位置を決定することで言語レベルが入力される。

【0042】以下、利用者が小学3年生であって、図4に示すように、画面206aでは「8～9」(年齢を表わす)のボタンが選択され、画面206bでは「小学3年」のボタンが選択され、画面206cでは小学3年の位置にカーソルが位置決めされた場合を例にして説明する。

【0043】なお、図4(a)～(c)に示す3つの言語レベル入力用の画面206a～206cは、何れか1つが機械翻訳装置に固定的に設けられていても良いし、何れかの形態を選択的に任意に用いることができるようになっていても良い。

【0044】まず、翻訳部104は、編集部103によって制御されているカーソルが、現在位置している原文を翻訳対象とする。ここでは、図3中に示す第5行目にある「Construction has begun on its three arenas, including the Olympic stadium, which after the Games will be trimmed by one third and converted into a major-league baseball park.」という

文の先頭にカーソルがあるものとして説明する。なお、翻訳部104における翻訳処理に必要な形態素解析、構文解析等の手法に関しては、既に様々なものが提案されているので詳細な説明を省略する。

【0045】翻訳部104は、翻訳処理の対象とする文章を、入力された利用者の言語レベルに照らし合わせて変換・生成を行なう。まず、翻訳部104は、言語レベルに関係なく共通する活用変化辞書105aや解析文法辞書105bを用いて、語尾等に変化のある単語・熟語をその原形に変換する処理や、文法の解析等を行なう。この言語レベルに関係なく共通である解析結果は、システム制御部102を介して文書記憶部107に保持させておく。さらに、翻訳部104は、図5に示すようなレベル別単語・熟語辞書105cの単語情報を参照しながら形態素解析を行なう。

【0046】レベル別単語・熟語辞書105cは、各見出し語(第一言語の単語)に対する訳語を、ひらがなで表記するか、あるいは漢字の訳語の場合であっても、振り仮名を付けるかなどの情報が記憶されており、レベル別に訳語が設定されている。レベル別単語・熟語辞書105cは、単語ごとに個別に、第一言語から第二言語に変換する変換処理に用いる。

【0047】次に、翻訳部104は、解析文法辞書105bを用いて、翻訳対象とする入力文を解析する。この場合、「Construction has begun on its three arenas, including the Olympic stadium.」の文については、例えば図6に示すように解析される。

【0048】次に、翻訳部104は、解析文法辞書105bを用いた解析結果を参照しながら、図7に示すような、レベル別変換文法辞書105dを用いて、利用者にとって理解しやすい構造に変換する。

【0049】具体的には、レベル別変換文法辞書105dには、埋め込み文を分割する規則1や受動態を能動態に変換する規則2など文の構造を単純にするための規則が蓄積されている。

【0050】本例の場合、規則1が適用され「Construction has begun on its three arenas, including the Olympic stadium.」という埋め込み文が「Construction has begun on its three arenas. They include the Olympic stadium.」のように分割される。

【0051】この後、翻訳部104は、レベル別単語・熟語辞書105c(図5)を参照して、変換処理によって単語ごとに個別に変換する。利用者が小学3年生であることを想定している場合、成人向け訳語として使用されている単語のうち、通常、4年生以上で学習する「含、競、技、建、設、縮、改、造」といった漢字を含む単語が原則として漢字表記されないようにレベル別単語・熟語辞書105cが作成されている。これらの漢字を含む単語は、ひらがなで表記させるか、あるいはそのまま残して、振り仮名を付与させることを示してい

る。また、小学3年レベルの漢字もまだ未習の可能性があるので、同様に振り仮名を付与させるようになっている。

【0052】一方、単語として小学校3年生には難解だと思われる「改造」といったような語は、図5に示すように「つくりかえる」というような和語を使った訳語にさせるようになっている。

【0053】次に、翻訳部104は、図8に示すようなレベル別生成文法辞書105eを用いて、言語レベルにあった訳文を生成するために、訳文中に挿入すべき読点の数を調整する。

【0054】例えば、「輸入された」という述語が、「アスバラガス[は]」、「食用[として]」、「明治時代[に]」、「アメリカ[から]」という名詞句によって修飾されている文の場合、小学3年の言語レベルでは、名詞句連続3以上のときに読点を挿入するという規則が、図8に示すレベル別生成文法辞書105eに記憶されているので、例えば「アスバラガスは、食用として明治時代にアメリカから輸入された。」のように読点が挿入される。

【0055】さらに、翻訳部104は、図9に示すようなレベル別形態素生成文法辞書105fに記憶された規則に従って、一般向けの訳文を言語レベルに応じて「である」調から「です・ます」調に変換したり、名詞に「お皿」の「お」に見られるような接頭語の「お」を付加したりする。本例では、利用者が小学3年の言語レベルであるので、訳文を「です・ます」調に変換することになる。

【0056】最後に、翻訳部104は、図10に示すような書式知識辞書105gに記憶された規則に従って、訳文に分かち書き、及び漢字に対して振り仮名を付与（初回のみ、あるいは該当する漢字の全て）して訳文を生成する。

【0057】翻訳部104によって翻訳対象となった原文に対して訳文が生成されると、編集部103は、この訳文を表示部106に表示させる（ステップA7）。図11には、表示部106の訳文表示用の画面306a、306bの一例を示している。

【0058】言語レベルが小学3年の場合には、図10の書式知識辞書105gに示すように、分かち書きをしない規則が記憶されているので、図11(a)に示すように、最終的な訳文である「その3か所のきょうぎ場で、けんせつの作業が始まりました。これはオリンピック・スタジアムをふくみます。ここは、オリンピックの後、3分の1に小さくして、メジャーリーグの野球場につくりかえます。」を表示部106の画面306aに表示させて利用者に提示する（なお、図11に示すように、漢字「所」には「しょ」、「始」には「はじ」、「野球」には「やきゅう」の振り仮名が付与されている）。

【0059】また、小学2年以下の言語レベルが指示されていた場合には、図11(b)に示すように、図11(a)に示す訳文に、さらに分かち書きが施されて表示される。

【0060】こうして、訳文の表示が完了すると、編集部103は、カーソルを翻訳が終了した文の次の文の先頭に移動させ（ステップA8）、ステップA3に戻る。ここで、さらに翻訳要求の指示があった場合には、システム制御部102は、各部を制御して、次の原文について同様に言語レベルに応じた翻訳を実行させる。

【0061】なお、第1実施形態の機械翻訳装置では、次のようなことも可能である。すなわち、最初に指定された言語レベルに合った翻訳を行なった後、利用者がレベルの変更を指示した場合には、翻訳部104は、言語レベルに関係なく共通して行なわれる処理の解析結果が文書記憶部107に保持されているので、この解析結果を利用して訳文の生成を行なう。すなわち、最初から翻訳をやり直す必要はなく、この解析結果のデータをもとに、言語レベルによって異なる部分の処理をレベル別単語・熟語辞書105c、レベル別変換文法辞書105d、レベル別生成文法辞書105e、レベル別形態素生成文法105f、書式知識辞書105gを用いて行えばよい。

（第2実施形態）次に、本発明の第2実施形態について説明する。

【0062】図12は第2実施形態に係わる機械翻訳装置の構成を示すブロック図である。図12に示すように、機械翻訳装置は、入力部101、システム制御部102、編集部103、翻訳部404、翻訳辞書部405、表示部106、及び文書記憶部107によって構成されている。なお、第1実施形態と同一の部分については同一の符号を付して説明を省略する。

【0063】翻訳部404は、システム制御部102で選択された第一言語による文章データに対して、システム制御部102で選択された翻訳辞書部405の辞書を参照しながら翻訳処理を行なう。

【0064】翻訳辞書部405は、翻訳部404が翻訳処理を行なう際に使用する各種辞書データを収容している。以下、翻訳辞書部405に含まれる辞書の内容について、図12の翻訳辞書部405を参照しながら説明する。

【0065】翻訳辞書部405は、第一言語から第二言語への標準的な翻訳を行なうための辞書と、同一の言語内、すなわち第二言語内で言語レベルに応じた書き替え（翻訳）を行なうための辞書の二段階構造になっている。

【0066】図12に示すように、標準的な翻訳を行なうための辞書には、語尾等に変化のある単語・熟語をその原形に変換する活用変化辞書405a、第一言語の単語・熟語に対応する第二言語の訳語がその品詞情報と共に

に記憶される一般単語・熟語辞書405h、第一言語を解析するための文法が記憶された一般解析文法辞書405b、第一言語の解析結果をもとに第二言語の構造に変換するために必要な情報が記憶された一般変換文法辞書405i、変換された第二言語の文の構造を決定する一般生成文法辞書405j、第二言語の訳語(単語)の語尾等の語形を変化させて一般的な翻訳文を完成させる一般形態素生成文法405k等を収容したものである。

【0067】一方、第二言語内で書き替えを行なうための辞書は、第二言語の単語・熟語がレベル別に分類されているレベル別単語・熟語辞書405c、第二言語を解析するための文法が記憶された第二言語解析文法辞書405l、同じ第二言語内で利用者の言語レベルに適した変換情報が記憶されたレベル別変換文法辞書405d、変換された第二言語の利用者向けに書き換えられた文の構造を決定するレベル別生成文法辞書405e、さらに語尾等の語形を変化させて利用者向けの翻訳文を完成させるレベル別形態素生成文法辞書405f、言語レベルに応じた特殊な書式に変換するために必要な書式知識辞書405g等を記憶したものである。

【0068】次に、第2実施形態における機械翻訳装置の動作について説明する。入力部101から入力された文章に対する翻訳処理の流れを、第一言語の英語による原文を、第二言語の日本語による訳文に翻訳する場合を例にして、図13のフローチャートを参照しながら説明する。

【0069】ここでは、具体例として、表示部106によって図3に示すように表示された英語の文章のうち、第5行目からの「Construction has begun on its three arenas, including the Olympic stadium, which after the Games will be trimmed by one third and converted into a major-league baseball park.」という文に対する翻訳の要求を受け、翻訳・表示する場合を例にして説明する。

【0070】なお、図13に示すステップB1～B5、B10の処理は、前述した第1実施形態における図2に示すフローチャートのステップA1～A5、A9の処理と同一であるので説明を省略する。また、翻訳レベルの入力の方法についても、第1実施形態と同様にして図4に示す言語レベル入力用の画面を用いて行なうものとする。

【0071】ところで、「翻訳キー」が押された場合(ステップB5)、システム制御部102は、翻訳要求が出されたことと見なし、以下のステップB6～B9の処理を各部において実行させる。まず、システム制御部102は、翻訳部404に対して、第一言語の原文を第二言語の訳文に翻訳する翻訳処理を、言語レベルに応じて実行させる。すなわち、第一言語の原文を翻訳することによって生成された第二言語の訳文について、言語レベルによって特定される表示形態となるように変換する。

【0072】ここでは、利用者が小学3年生の言語レベルを指示した場合を例にして説明する。まず、翻訳部104は、編集部103によって制御されているカーソルが、現在位置している原文を翻訳対象とする。ここでは、図3中に示す第5行目にある「Construction has begun on its three arenas, including the Olympic stadium, which after the Games will be trimmed by one third and converted into a major-league baseball park.」という文の先頭にカーソルがあるものとして説明する。なお、翻訳部404における翻訳処理に必要な形態素解析、構文解析等の手法に関しては、既に様々なものが提案されているので詳細な説明を省略する。

【0073】翻訳部404は、翻訳辞書部405の標準的な翻訳を行なうための辞書を参照して、第一言語から第二言語への通常の翻訳を行なう(ステップB6)。ステップB6における翻訳結果としては、例えば「オリンピック・スタジアムを含む、その3か所の競技場で建設作業が始まった。これは、オリンピックの後は3分の1に縮小され、メジャーリーグの野球場に改造される。」のような訳文が得られる。

【0074】次に、翻訳部404は、ステップB6における翻訳処理によって得られた、言語レベルに関係のない一般向け訳文について、さらに利用者の言語レベルに照らし合わせて解析・変換・書き替えを行なう。すなわち、翻訳部404は、第二言語内で、指示された言語レベルに応じた翻訳を行なう(ステップB7)。

【0075】まず、翻訳部404は、図14に示すようなレベル別単語・熟語辞書405cを参照し、訳文で用いられている各単語について、書き替えが必要かどうかを調べる。

【0076】図14に示すレベル別単語・熟語辞書405cにおいて、「0」は変更を加える必要のないことを示しており、それ以外は該当する単語に変えることを示している。利用者が小学3年であることを想定している場合、4年生以上で学習する「含、競、技、建、設、縮、改、造」といった漢字を含む単語が原則として漢字表記されないようにレベル別単語・熟語辞書405cが作成されている。これらの漢字を含む単語は、ひらがなに変換するか、あるいはそのまま残して、振り仮名を付与させることを示している。

【0077】また、小学3年レベルの漢字もまだ未習の可能性があるので、同様に振り仮名を付与させるようになっている。一方、単語として小学校3年生には難解だと思われる「改造」といったような語は、図14にあるように「つくりかえる」というように和語を使って言い換える必要があることを示している。

【0078】次に、翻訳部404は、第二言語解析文法辞書405lを用い、解析文法辞書405bを用いて入力文を解析した場合と同様の手法で、一般の翻訳処理で

生成された訳文を解析する。この訳文の場合、例えば図15のように解析される。

【0079】次に、翻訳部404は、解析結果を参照しながら、レベル別変換文法辞書405d(図7に示すレベル別変換文法辞書105dと同一)を用いて、現状の訳文が利用者にとって理解しづらいか否かを言語レベルに基づいて判断する。

【0080】レベル別変換文法辞書405dには、埋め込み文を分割する規則1や受動態を能動態に変換する規則2など文の構造を単純にするための規則が蓄積されている。本例の場合では、規則1が適用され、「オリンピック・スタジアムを含む、その3か所の競技場で建設作業が始まった。」の部分が「その3か所の競技場で建設作業が始まった。オリンピック・スタジアムを含む。」のように分割される。

【0081】次に、翻訳部404は、レベル別生成文法辞書405e(図8に示すレベル別生成文法辞書105eと同一)を用いて、言語レベルにあった訳文を生成するために、一般向けの訳文にさらに読点を追加して訳文を生成する。

【0082】レベル別生成文法辞書405eに記憶された規則では、例えば「アスパラガスは食用として明治時代にアメリカから輸入された。」という訳文の場合、小学3年生の言語レベルでは、名詞句連続3以上のときに読点を挿入するという規則があるため、例えば「アスパラガスは、食用として明治時代にアメリカから輸入された。」のように読点が挿入される。

【0083】さらに、翻訳部404は、レベル別形態素生成文法405f(図9に示すレベル別形態素生成文法辞書105fと同一)を用いて、一般向けの訳文を言語レベルに応じて「である」調から「です・ます」調に変換したり、名詞に「お皿」の「お」に見られるような接頭語の「お」を付加したりする。本例では、利用者が小学3年生であるので、「です・ます」調に変換することになる。

【0084】最後に、翻訳部404は、書式知識辞書405k(図10に示す書式知識辞書105gと同一)に記憶された規則に従って、訳文に分ち書き、及び漢字に対して振り仮名を付与(初回のみ、あるいは該当する漢字の全て)して訳文を生成する。

【0085】翻訳部404によって翻訳対象となった原文に対して訳文が生成されると、編集部103は、この訳文を第1実施形態の場合と同様にして、図11に示すように表示部106に表示させる(ステップB8)。

【0086】こうして、訳文の表示が完了すると、編集部103は、カーソルを翻訳が終了した文の次の文の先頭に移動させ(ステップB9)、ステップB3に戻る。このようにして、第1実施形態及び第2実施形態において説明した翻訳方法によれば、利用者の言語レベルに応じて利用者が通常接している文章形態で原文に対する訳

文を生成し、提示することができるので、母国語以外の文章を人手を借りることなく容易に理解し、情報を得ることができる人々の層が増加する。

【0087】なお、本発明は、前述した実施形態に限定されず、要旨を変更しない範囲で適宜変形して実施できる。まず、前述した実施形態では英語の文章を日本語に翻訳する場合を例にして説明したが、逆に日本語の文章を英語の文章に翻訳する場合にも応用することができる。この場合、訳文の難易度はラテン語系統の語とゲルマン語(古英語)系統の使い分けで決められるであろう。例えば、「庭園は夏には花一面になる。」の訳は、標準的には「There is a profusion of flowers in the garden in summer.」となるが、読み手が例えば英語圏の小学一年の場合、「There is a lot of flowers in the garden in summer.」となる。

【0088】また、同じ英語から日本語への翻訳であっても、利用者等が高齢者の場合は最近のカタカナ語を極力使わないように訳文を書き換えるといったことも考えられる。さらに、ステップA7、B8では、最終的な訳文のみを表示するとしたが、途中で得られる一般向けの訳文をあわせて表示することも可能である。

【0089】さらに、訳文を視覚的に提示するために表示しているが、必ずしも表示させる必要はなく、例えば音声合成を用いて出力することも可能である。これに加え、利用者がとくに低学年の場合は関連する絵をあわせて表示し、理解の一助とすることも考えられる。

【0090】

【発明の効果】以上詳述したように本発明によれば、利用者の言語レベルに応じて利用者が通常接している文章形態で原文に対する訳文を生成し提示するので、利用者は言語能力とは関係なく母国語以外の文章を人手を借りることなく容易に理解することができる。従って、訳文から情報を得ることができる人々の層を増加させることができる。

【図面の簡単な説明】

【図1】本発明の第1実施形態に係わる機械翻訳装置の構成を示すブロック図。

【図2】第1実施形態における英語を日本語に翻訳する場合の翻訳処理の流れを示すフローチャート。

【図3】第1実施形態の翻訳処理を説明するための文章の具体例を示す図。

【図4】第1実施形態における言語レベル入力用の画面の一例を示す図。

【図5】図1中に示すレベル別単語・熟語辞書105cの一例を示す図。

【図6】第1実施形態を説明するための文章の解析結果の一例を示す図。

【図7】第1実施形態のレベル別変換文法辞書105dの一例を示す図。

【図8】第1実施形態のレベル別生成文法辞書105e

の一例を示す図。

【図9】第1実施形態のレベル別形態素生成文法辞書105fの一例を示す図。

【図10】第1実施形態の書式知識辞書105gの一例を示す図。

【図11】第1実施形態に用いられる表示部106の訳文表示用の画面のレイアウトの一例を示す図。

【図12】本発明の第2実施形態に係わる機械翻訳装置の構成を示すブロック図。

【図13】第2実施形態における英語を日本語に翻訳する場合の翻訳処理の流れを示すフローチャート。

【図14】第2実施形態のレベル別単語・熟語辞書405cの一例を示す図。

【図15】第2実施形態を説明するための文章の解析結果の一例を示す図。

【符号の説明】

101…入力部

102…システム制御部

103…編集部

104, 404…翻訳部

105, 405…翻訳辞書部

105a, 405a…活用変化辞書

105b, 405b…解析文法辞書

105c, 405c…レベル別単語・熟語辞書

105d, 405d…レベル別変換文法辞書

105e, 405e…レベル別生成文法辞書

105f, 405f…レベル別形態素生成文法辞書

105g, 405g…書式知識辞書

106…表示部

107…文書記憶部

405h…一般単語・熟語辞書

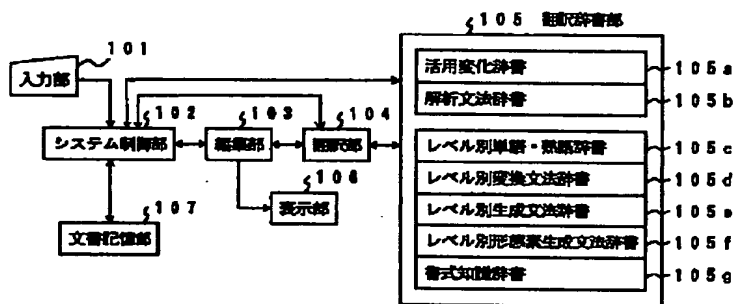
405i…一般変換文法辞書

405j…一般生成文法辞書

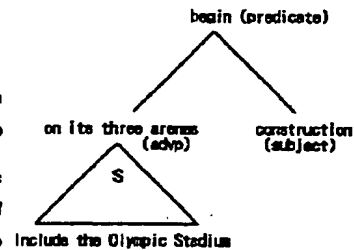
405k…一般形態素生成文法辞書

405l…第二言語解析文法辞書

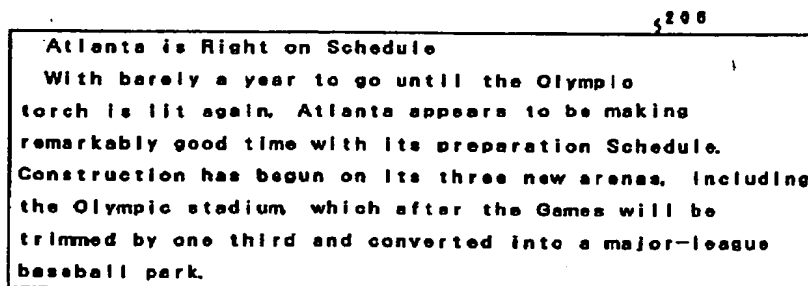
【図1】



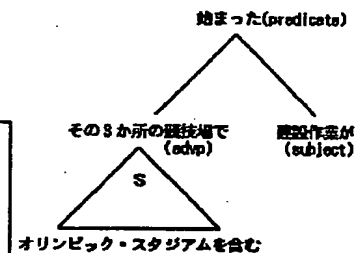
【図6】



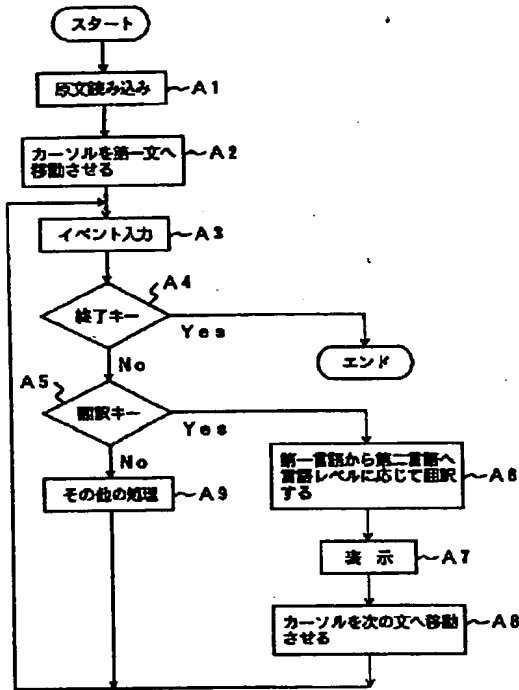
【図3】



【図15】



【図2】



【図11】

(a)

§ 308a

その3か所のきょうぎ場で、けんせつの作業が<sup>BC</sup>始まりました。これはオリンピック・スタジアムをふくみます。ここは、オリンピックの後、3分の1に小さくして、メジャーリーグの<sup>野球場</sup>につくりかえます。

(b)

§ 308b

その3か所のきょうぎ場で、けんせつの作業が<sup>BC</sup>始まりました。これは、オリンピック・スタジアムをふくみます。ここは、オリンピックの後、3分の1に小さくして、メジャーリーグの<sup>野球場</sup>につくりかえます。

【図4】

(a) § 208a

〈利用者年齢〉

1~3	4~6	6~7	8~9	10~12	13~15	15~18	20代	30~40代	50以上
-----	-----	-----	-----	-------	-------	-------	-----	--------	------

(b) § 208b

〈利用者学級〉

幼児	小学1年	小学2年	小学3年	小学4年	小学5年	小学6年	中学	高校	大学
----	------	------	------	------	------	------	----	----	----

(c) § 208c

〈利用者学級〉

幼児	小学1年	小学2年	小学3年	小学4年	小学5年	小学6年	中学1年	中学2年	中学3年	高校1年	高校2年	高校3年	大学
----	------	------	------	------	------	------	------	------	------	------	------	------	----

5105c レベル別単語・熟語辞書

英辞レベル		高小中							
単語		校	学	0	5	4	3	2	1
begin	verb	始まる					はじまる		はじまる
include	verb	含む	い	ふくむ					
trial	verb	試小する			いししょう 試小する	ししょう する	小さくする		
convert	verb	変換する		いへん 変換	作り変える	作りかえる	つくりかえる		
construction	noun	建設			けんせつ 建設	けんせつ 建設	けんせつ	建てる	たてる
arena	noun	競技場				きょうぎ 競技場	きょうぎ きょうぎ場		きょうぎ じょう
baseball park	noun	野球場				やきゅう 野球場		やきゅう 野球場	やきゅう じょう

5105d レベル別変換文法辞書

事項 言語 レベル	高 校	中 学	小 5	小 6	小 4	小 3	小 2	小 1
埋め込み	0				規則1			
受動語	0			規則2				
・規則1 ((	) S1	) S2	→	(	) S2	(	) S1	
規則2 (	) NP	(	) VP	→	(	) VP	(	) NP
			受身			能動		

5105 ● レベル別生成文法辞書

事項 書面 レベル	高 校	中 学	小 5	小 5	小 4	小 3	小 2	小 1
読み挿入箇所	名詞句連続 6 以上		名詞句連続 6 以上		名詞句 連続 4 以上	名詞句 連続 3 以上	名詞句連続 2 以上	
	動詞句連続 3 以上				動詞句連続 2 以上			

【図9】

§105f レベル別形態素生成文法辞書

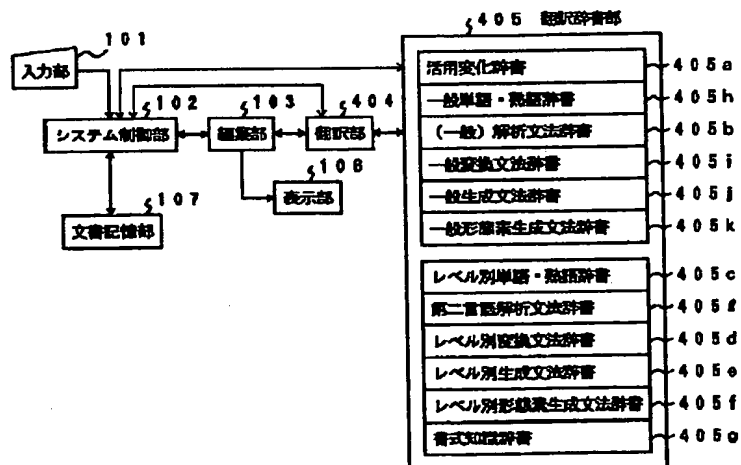
事項 言語 レベル	高 校	中 学	小 6	小 5	小 4	小 3	小 2	小 1
です・ます類 ／で・ある類	で・ある類					です・ます類		
接頭語「お」 の付与	付与しない						付与する	

【図10】

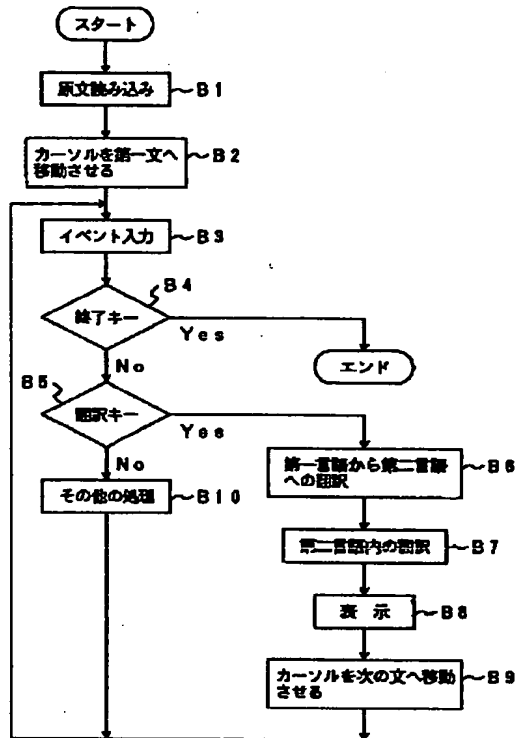
§105g 書式知識辞書

言語 レベル	高 校	中 学	小 6	小 5	小 4	小 3	小 2	小 1
事項	被	学	6	5	4	3	2	1
分ち書き	しない						する	
振り仮名	初出のみ					該当の漢字すべて		

【図12】



【図13】



【図14】

4050 レベル別単語 - 単語辞書

単語	高	中	小	小	小	小	小	小
	校	学	6	5	4	3	2	1
含む	助詞	0	ふくむ	ふくむ				
始まる	助詞	0				はじまる	はじまる	
縮減する	サ変	0		むく	ひんせつ	けんせつ	と縮める	たてる
作成する	サ変	0					ひんせつ	さげよう
縮小する	サ変	0		しゅくしょう	しゅくしょう	小さくする		
改定する	サ変	0		かいせい	作り変える	作りかえる	つくりかえる	
競技場	名詞	0			きょうぎ	きょうぎ	きょうぎ	きょうぎ
野球場	名詞	0			やきゅう	やきゅう	やきゅう	やきゅう
後	名詞	0						あと

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**